

Requested Patent: JP8297675A
Title: TRANSLATION SUPPORTING DEVICE ;
Abstracted Patent: JP8297675 ;
Publication Date: 1996-11-12 ;
Inventor(s): KAIKOU TATSUYA; YAMAMOTO HIDEKI ;
Applicant(s): OKI ELECTRIC IND CO LTD ;
Application Number: JP19950194134 19950707 ;
Priority Number(s): ;
IPC Classification: G06F17/28 ;

Equivalents:

ABSTRACT:

PURPOSE: To shorten the time for translation including final correction by processing the coincident parts of a former source sentence and a new source sentence while utilizing a former translated sentence as it is, and processing completely different parts while using machine translation and translation check, etc.

CONSTITUTION: A document comparing part 3 reads a former source sentence file 1 and a new source sentence file 2. Then, the difference between the former source sentence and the new source sentence is investigated and difference information is prepared. That difference information is read block by block and it is judged whether there is any difference between the correspondent former and new source sentences or not. When it is judged that there is no difference, a translated sentence corresponding to that full sentence is found from a former translated sentence file 11. Then, that translated sentence is added to a new translated sentence file 12 as it is. When it is judged there is any difference, on the other hand, a differential value is calculated. Then, it is compared with a difference threshold value. When the differential rate is larger than the difference threshold value, the source sentence is mechanically translated. Then, the translation is checked by a user while displaying the source sentence and the translated sentence.

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-297675

(43) 公開日 平成8年(1996)11月12日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/28		8420-5L	G 0 6 F 15/38	P
		8420-5L		T

審査請求 未請求 請求項の数9 F D (全 13 頁)

(21) 出願番号 特願平7-194134

(22) 出願日 平成7年(1995)7月7日

(31) 優先権主張番号 特願平7-64910

(32) 優先日 平7(1995)2月28日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72) 発明者 介弘 達哉

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72) 発明者 山本 秀樹

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

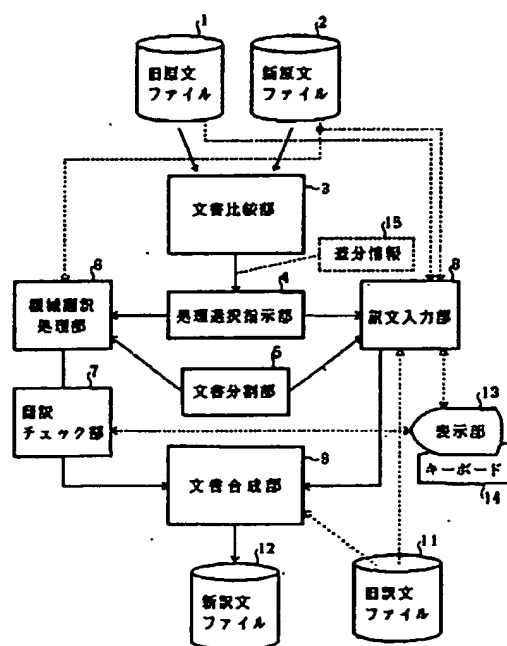
(74) 代理人 弁理士 佐藤 幸男 (外1名)

(54) 【発明の名称】 翻訳支援装置

(57) 【要約】

【目的】 既に機械翻訳と翻訳チェックの終了した旧原文の一部を変更した新原文について翻訳を行う場合に、その翻訳処理を効率よく行う。

【構成】 文書比較部3で旧原文と新原文とを比較する。対応する部分についてその差が著しい場合には機械翻訳と翻訳チェックを経て新訳文を得る。差分が少ない場合には、旧訳文及び、どの部分が相違するか相違する部分を表示して、訳文入力部8による新たな訳文の入力を促す。順にこの作業を繰り返し、旧訳文で利用できる部分をフルに活用する。



本発明の翻訳支援装置ブロック図

1

【特許請求の範囲】

【請求項1】 既に機械翻訳処理され、翻訳チェックがされた旧原文とその翻訳を含む旧訳文を使用して、前記旧原文の一部を変更した新原文の翻訳処理を実行するものにおいて、

前記旧原文と新原文との相違する部分を抽出して差分情報を生成する文書比較部と、

表示された差分情報の内容に基づいて判断される前記旧訳文の利用方法を指示する処理選択指示部と、

旧訳文を利用しないとき、新原文の該当箇所を機械翻訳する機械翻訳処理部と、

旧訳文を利用するとき、旧原文と新原文と旧訳文と前記差分情報とを表示部に表示して、新訳文の入力を受け付ける訳文入力部と、

前記旧訳文の該当する部分を、前記新訳文に置き換える文書合成部とを備えたことを特徴とする翻訳支援装置。

【請求項2】 旧原文と新原文の対応する部分を取り出して、それぞれ、単位量のデータごとにその有無を比較し、その文を構成する全てのデータ数に対する他方の文に無いデータ数の割合を差分率としたとき、新原文と旧原文の差分率の大きい方を前記差分情報に含め、処理選択指示部は、

予め設定した差分率の閾値と前記差分情報中の差分値とを比較して、

差分情報中の差分値が閾値よりも大きい場合に、旧訳文を利用しないで機械翻訳処理を実行するように指示することを特徴とする請求項1記載の翻訳支援装置。

【請求項3】 単位量のデータを、新旧原文を構成する単語とすることを特徴とする請求項2記載の翻訳支援装置。

【請求項4】 訳文入力部は、前記差分情報に基づいて、一方の原文に含まれており他方の原文に含まれないデータに、各原文中で他のデータと区別できる識別表示を付して、旧原文と新原文と旧訳文とを表示するように表示部を制御することを特徴とする請求項1～3に記載の翻訳支援装置。

【請求項5】 旧原文と旧訳文とを構成する各文を比較して、旧原文の各文の翻訳文は旧訳文のどの文に相当するかの対応付けを行う文書対応付け部を備えたことを特徴とする請求項1記載の翻訳支援装置。

【請求項6】 旧原文と新原文とを比較して得た差分情報と、旧原文と新原文と旧訳文及び新訳文を、比較対象の文の前後の文を含めて示すウインドウとを表示する差分表示部を備えたことを特徴とする請求項1記載の翻訳支援装置。

【請求項7】 旧原文と新原文と旧訳文及び新訳文を、比較対象の文の前後の文を含めて示すとともに、旧訳文の任意の部分を新訳文編集用のウインドウに複写して、修正編集処理を可能にする翻訳チェック表示部を備えたことを特徴とする請求項1記載の翻訳支援装置。

2

【請求項8】 差分率の閾値を任意の値に選択して設定する差分値選択部を備えたことを特徴とする請求項2記載の翻訳支援装置。

【請求項9】 旧原文と新原文とが同一の部分は、旧訳文を、その文書情報を含めてそっくり新原文に複写する訳文変換手段を備えたことを特徴とする請求項1記載の翻訳支援装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、文書の機械翻訳装置等と共に翻訳作業を支援するための翻訳支援装置に関する。

【0002】

【従来の技術】 自然言語で作成された文章を例えば英語から日本語に翻訳する場合に、コンピュータにより自動的にその処理を実行する機械翻訳装置は、既に多くの分野に利用されるようになってきた。その翻訳の質向上のために、例えば特公平3-51020号公報等に各種の工夫が紹介されている。この文献には、単語の情報を保持した辞書と原文の構文解析結果と訳文の生成を行う文法知識とを用いて、原文を目標言語の文に翻訳する機械翻訳装置の構成等が紹介されている。

【0003】 一般に、自然言語の曖昧さのために、機械翻訳装置の出力する訳文はそのままでは使用できない場合がある。この場合には通常、人手により翻訳チェックを実施して最終的な訳文を生成する。この他、機械翻訳装置に備わっているユーザごとに単語を登録する機能を用いて専門用語等の単語を登録してから、再度機械翻訳を実施するといった方法がある。人手による翻訳チェックと単語辞書を更新しながら行う機械翻訳とを組み合わせ、更に質の高い訳文を得ることも行われている。

【0004】

【発明が解決しようとする課題】 ところで、従来の機械翻訳を用いた翻訳支援装置には次のような解決すべき課題があった。従来の機械翻訳装置を用いて、過去に翻訳した文章の一部だけを変更した文章を翻訳しようとする場合、新たな原文の翻訳と同様に、初めから機械翻訳を実施し、その後翻訳チェックを行う方法がある。また、人手により新たな原文と古い元の原文との相違点を見つけ出し、その部分を翻訳して元の訳文を修正するといった方法がある。

【0005】 しかしながら、原文の量が膨大になった場合には、原文と訳文の相違点を見つける作業は容易でない。また、既に翻訳チェックまで完了した元の訳文を放棄して、再び新たな原文の機械翻訳結果を得てその翻訳チェックを実施するのは、多くの部分で無駄な作業時間を費やすことになる。しかも、翻訳後の訳文の文書フォーマットを整える作業も、改めて全文について実施するのは二度手間になり、無駄が多い。

【0006】

3

【課題を解決するための手段】本発明は上記の点を解決するため次の構成を採用する。既に機械翻訳処理され、翻訳チェックがされた旧原文とその翻訳を含む旧訳文を使用して、旧原文の一部を変更した新原文の翻訳処理を実行するものにおいて、旧原文と新原文との相違する部分を抽出して差分情報を生成する文書比較部と、表示された差分情報の内容に基づいて判断される旧訳文の利用方法を指示する処理選択指示部と、旧訳文を利用しないとき、新原文の該当箇所を機械翻訳する機械翻訳処理部と、旧訳文を利用するとき、旧原文と新原文と旧訳文と差分情報とを表示部に表示して、新訳文の入力を受け付ける訳文入力部と、旧訳文の該当する部分を、新訳文に置き換える文書合成部とを備える。

【0007】なお、旧原文と新原文の対応する部分を取り出して、それぞれ、単位量のデータごとにその有無を比較し、その文を構成する全てのデータ数に対する他方の文に無いデータ数の割合を差分率としたとき、新原文と旧原文の差分率の大きい方を差分情報に含め、処理選択指示部は、予め設定した差分率の閾値と差分情報中の差分値とを比較して、差分情報中の差分値が閾値よりも大きい場合に、旧訳文を利用しないで機械翻訳処理を実行するように指示するとよい。

【0008】また、単位量のデータを、新旧原文を構成する単語とすることができる。更に、訳文入力部は、差分情報に基づいて一方の原文に含まれており他方の原文に含まれないデータに、各原文中で他のデータと区別できる識別表示を付して、旧原文と新原文と旧訳文とを表示するように表示部を制御するとよい。

【0009】更に、旧原文と旧訳文とを構成する各文を比較して、旧原文の各文の翻訳文は旧訳文のどの文に相当するかの対応付けを行う文書対応付け部を備えるといふ。また、旧原文と新原文とを比較して得た差分情報と、旧原文と新原文と旧訳文及び新訳文を、比較対象の文の前後の文を含めて示すウインドウとを表示する差分表示部を備えてもよい。また、旧原文と新原文と旧訳文及び新訳文を、比較対象の文の前後の文を含めて示すとともに、旧訳文の任意の部分を新訳文編集用のウインドウに複写して、修正編集処理を可能にする翻訳チェック表示部を備えるといふ。

【0010】また、差分率の閾値を任意の値に選択して設定する差分値選択部を備えるといふ。また、旧原文と新原文とが同一の部分は、旧訳文を、その文書情報を含めてそっくり新原文に複写する訳文変換手段を備えるといふ。

【0011】

【作用】まず予め、旧原文と旧訳文とを用意し、文の対応付けをしておく。次に文書比較部で旧原文と新原文とを比較する。対応する部分についてその差が著しい場合や新たな挿入文等の場合には機械翻訳と翻訳チェックを経て新訳文を得る。差分が少ない場合には、旧原文や新

4

原文と共に、旧訳文及び、どの部分が相違するか相違する部分を表示して、訳文入力部による新たな訳文の入力を促す。差が全く無ければ対応する旧訳文をそっくりコピーする。順にこの作業を繰り返し、旧訳文で利用できない部分をフルに活用する。このように、旧原文と新原文が同一の部分は全く同一の翻訳結果が得られるので、翻訳品質が安定する。

【0012】旧原文と新原文との相違する部分を抽出する処理は自動的に行われる。その具体的な方法として、旧原文を構成する全てのデータ数に対応する新原文に無いデータ数の割合を差分率として求める。新原文についても同様の差分率を求める。いずれか多い方を選択すると、比較対象となった文の相違が定量的に把握できる。これを任意に設定した閾値と比較し、差分の大きいものは旧訳文を放棄して機械翻訳を行い、翻訳チェックにより新たな訳文を得る。その他の場合には可能な限り旧訳文を利用し、その修正等によって新訳文を得る。英語の場合、差分率演算に用いられる単位量のデータは単語が適切である。日本語の場合には文節になる。このような文章の切り分けは構文解析の手法による。訳文入力部で旧原文、新原文、旧訳文、差分情報等を表示するのは、訳文入力部の便宜のためである。差分情報の表示方法として、新原文と旧原文の相違部分にアンダーラインや色分け等の表示を行えば、比較翻訳作業が容易になる。旧訳文をそっくりコピーするときは、インデントや空行等の文書情報も含めてコピーすれば、文書フォーマットの校正作業も省略できる。

【0013】

【実施例】以下、本発明を図の実施例を用いて詳細に説明する。図1は、本発明の翻訳支援装置実施例を示すブロック図である。この装置は、旧原文ファイル1と新原文ファイル2とを比較しながら、必要に応じて旧訳文ファイル11を利用し、新訳文ファイル12を生成するための装置である。ここでは、文書比較部3、処理選択指示部4、文書分割部5、機械翻訳処理部6、翻訳チェック部7、訳文入力部8、文書合成部9及び表示部13とキーボード14が設けられている。

【0014】旧原文ファイル1は記憶装置上に設定され、既に機械翻訳され翻訳チェックが終了した例えば英文の文書が格納されている。新原文ファイル2も同様に記憶装置上に設定され、旧原文の一部を変更して構成された英文等から成る。図2と図3に、このような旧原文ファイルと新原文ファイルの例説明図を示す。各図の左側には行番号を示し、1行が2段に亘って記載されている場合には左側の行番号を空白にしている。例えば、旧原文ファイルは図に示すような英文から構成され、1行目前段、1行目後段～3行目、4行目前段、4行目後段の4個の文から構成される。図3に示した新原文は図2に示した旧原文の一部を修正しあるいは追加して作成された新たな英文である。これは、図の左側に示すよう

5

に、1行目前段、1行目後段～5行目、6行目の前段と後段の4個の文から構成されている。

【0015】図4に、旧訳文ファイルの例説明図を示す。図1に示した旧訳文ファイル11には、図2に示した旧原文を機械翻訳し、更に人手によって翻訳チェックを完了した日本語が格納されている。また、これは訳文の文書フォーマットの校正も完全に終了したものである。これは、図4に示すように、旧原文の4個の文に対応する4個の訳文から構成されている。

【0016】図1に示す文書比較部3は、旧原文と新原文との相違する部分を抽出して、差分情報を生成する部分である。即ち、図2と図3に示す旧原文及び新原文の1行目を比較した場合に、その丁度中程にある旧原文の“outputs”が“displays”に置き代わっている。このような相違部分の抽出を構文解析と単語比較によって実施する。そして、その結果を文書比較部3は差分情報15として出力する。なお、このような処理は、既によく知られた技術により実施でき、例えばdiffや、マイクロソフト社によるMS-DOS標準のfcや、Free Software Foundation, Inc. のwdiff等によって実現可能である。

【0017】図5には、その差分情報の例説明図を示す。ここでは、新原文と旧原文の相違部分を□でくくって他の部分と区別できるように表示している。□の中の一記号で挟まれた単語は新原文において削除された単語である。また□の中で+記号に挟まれた単語は新原文において新たに追加された単語である。行頭の!の記号は旧原文と新原文において差分が生じている行を示し、行頭の+記号は新原文で新たに追加された行を示す。行頭の=記号は新旧の原文において差分のなかった即ち一致する行を示す。文書比較部3はこのような差分情報を更に加工し、図6に示すような差分率を演算する。

【0018】この差分率というのは、本発明において、コンピュータ処理に適する手法として新たに設定されたもので、対応する旧原文と新原文に着目した場合に、その差を定量的に表すための数値である。その場合に、まず単位量のデータを設定する。この単位量のデータは英文の場合、構文解析等によって認識された単語でよい。もちろん、単に文字1個1個を単位量のデータと取り扱っても比較は可能である。例えば、図2と図3に示した旧原文及び新原文の1行目を比較した場合に、旧原文のワード数は14である。その差分ワード数は1となる。新原文についても全く同様で、ワード数は14、差分ワード数は1となる。

【0019】差分率は、比較対象部分を構成する全てのデータ数、即ち全てのワード数に対する他方の部分にないデータ数をパーセントで表したものとす。この場合、旧原文の差分率は $1 \div 14$ で7パーセント、新原文の差分率は同様に7パーセントである。従って、いずれ

6

の差分率をとっても結論としては7パーセントという値が得られる。一方、2行目についても同様に差分率をとったとき、旧原文も新原文も同じ $1 \div 3$ で33パーセントとなっている。一方、3行目は旧原文のワード数が9、差分ワード数が3、従って、差分率は33パーセントとなる。これに対して新原文のワード数は10、差分ワード数は4であるから、差分率は40パーセントとなる。

【0020】この場合に、旧原文と新原文の差分率を比較し、多い方の差分率を選択して両者の比較結果とする。従って、結果として3行目の差分率は40パーセントとなる。この差分率は、先に説明したように、比較する旧原文と新原文との相違を定量的に示すための数値である。従って、この数値が大きいほど違いが著しく、旧原文を翻訳した旧訳文は殆ど利用できないことになる。ここで、どの程度の差分率で旧訳文を利用するかを差分率の閾値として設定しておく。この閾値が、例えば50パーセントであれば、差分率50パーセントまでは旧訳文を利用した処理が実行される。一方、差分率50パーセント以上であれば旧訳文は放棄し、通常の機械翻訳処理を新原文に対して実行する。

【0021】差分情報の4行目と5行目が新原文で新しく追加されたものなので、差分率は100パーセントとなる。従って、新原文は機械翻訳処理部6に渡されて翻訳が行われる。差分情報の6行目は新旧原文とも差が無いので、旧訳文がそのまま利用されることになる。

【0022】図1に示す処理選択指示部4はこのような閾値を保持し、機械翻訳処理部6を起動するか訳文入力部8を起動するかを選択する、選択指示制御のための回路から構成される。文書分割部5は、旧原文ファイル1や新原文ファイル2の対応する文を切り出して、機械翻訳処理部6や訳文入力部8に転送する文書処理を実行する部分である。なお、このような文の切り分け、単語の切り分け、構文解析等は従来の自然言語文書処理のための装置に広く利用されている手法をそのまま使用すればよい。翻訳チェック部7は機械翻訳処理部6によって従来よく知られた方法で機械翻訳された結果を出力し、表示部13に表示する制御を行う部分である。表示部13に表示された機械翻訳の結果は翻訳者によって参照され、キーボード14を用いて誤訳や不適切な訳を行った部分の修正が実施される。

【0023】図7に、翻訳チェック時の表示画面例説明図を示す。この翻訳チェック処理そのものは従来の機械翻訳装置を用いた処理で実施されていたものであるが、この図に示すように、画面20には、原文表示ウィンドウ21と、訳文表示ウィンドウ22と、終了ボタン23と、スクロールバー24、25が表示されている。原文表示ウィンドウ21には、図に示すように、機械翻訳対象となった旧原文と新原文の差分の著しい行が1行表示される。訳文表示ウィンドウには図1に示す機械翻訳処

7

理部6によって自動的に翻訳処理された訳文が1行表示される。

【0024】なお、スクロールバー24は、原文や訳文の前後の関係を知りたい場合に、直前の文や直後の文を表示するために操作される。また、訳文表示ウィンドウ22に表示された訳文は翻訳者がキーボード等を用いて適宜修正等を加えて翻訳チェックを実施する。翻訳チェックが終了した場合、終了ボタン23が押されると、これが新訳文となって、図1に示す文書合成部9に転送される。図1に示す訳文入力部8は、旧訳文をそのまま使用するかあるいは旧訳文を利用して新たな訳文を生成する場合に使用される。その場合の操作画面は、図1に示す表示部13に表示され、翻訳者はキーボード14等を用いて新訳文の入力処理を実行する。

【0025】図8に、訳文入力時の表示画面例説明図を示す。この画面30には、旧原文表示ウィンドウ31と、新原文表示ウィンドウ32と、旧訳文表示ウィンドウ33と、新訳文表示ウィンドウ34が設けられている。これらにそれぞれ旧原文、新原文、旧訳文が表示され、新訳文表示ウィンドウ34には新訳文の表示と入力20が実行される。コピーボタン35は、これを押すことによって旧訳文表示ウィンドウ33に表示された訳文を新訳文表示ウィンドウ34にそのままコピーする処理を実行するボタンである。終了ボタン36は作業終了後に押すボタンである。スクロールバー37、38、39、40は、図7を用いて説明したものと同様、直前あるいは直後の文を参照したい場合に操作するためのものである。

【0026】なお、図8において、旧原文表示ウィンドウ31と新原文表示ウィンドウ32を眺めると、それぞれ互いに相違する部分にはアンダーラインが付されている。即ち、ここでは差分情報として、一方の原文に含まれており他方の原文に含まれない単語に他の単語と区別できるような識別表示を施している。これによって、翻訳者が相違部分を認識し確認しながら旧訳文を手直しして新たな訳文を得ることができる。なお、この識別表示はアンダーラインの他、その部分の単語をブリンクさせたり、高輝度表示したり、他の部分と異なる色で表示するといった方法が考えられる。図1に示す文書合成部9は、翻訳チェック部7あるいは訳文入力部8から転送された訳文を新訳文ファイル12に書き込む処理を実行する部分で、必要に応じて旧訳文ファイル11から該当する訳文を取り出し、新訳文ファイル12に転送する処理を実行する。

【0027】以上の構成の本発明の装置は次のように動作する。図9は、本発明の装置による処理フローチャートである。まず、ステップS1において、文書比較部3が旧原文ファイル1と新原文ファイル2を読み込む。そして、ステップS2において、旧原文と新原文の差分を調べ、差分情報を作成する。この差分情報は、例えば図

8

5に示したようなりスト状に生成される。ステップS3の次にあるステップS4においては、その差分情報から1ブロックずつ読み込みを行い、対応する旧原文と新原文との間に差分があるかどうかをステップS5で判断する。ステップS3は差分情報の全てのブロックについてこのような処理が済んだかどうかを判断する部分で、処理が終了していれば全体の処理が完了となる。

【0028】ステップS5において差分がないと判断された場合には、ステップS9に移り、旧訳文ファイルからその全文に対応する訳文を見つける。そして、それをそのまま新訳文ファイルに追加するようステップS8の処理に進む。一方、ステップS5において差分があると判断された場合には、ステップS6に移り、先に図6を用いて説明したような差分値を演算する。そして、差分閾値と比較する。差分率が差分閾値よりも大きい場合、あるいはこの例では両者が等しい場合には、ステップS6からステップS10に移り、原文を機械で翻訳する。そして、ステップS11において、原文及び訳文を表示しながら翻訳者であるユーザに翻訳チェックをさせる。そのチェック結果は、ステップS8において、新訳文ファイルに追加される。一方、ステップS6において、差分率が差分閾値よりも小さいと判断された場合には、ステップS7に移り、旧原文、新原文、差分情報、旧訳文を表示し、ユーザに新訳文を入力させる。こうして得られた新訳文がステップS8において、新訳文ファイルに追加される。

【0029】図10に、本発明の別の実施例による装置のブロック図を示す。この装置は、図1に示した装置に加えて、文書対応付け部51、単語辞書52、訳文変換手段53、翻訳チェック表示部54、差分表示部55及び対訳チェック表示部56を追加している。なお、図1に示したものと同一部分には同一符号を付して重複する説明は省略する。ここで、文書対応付け部51は、旧原文ファイル1と旧訳文ファイル11とを調べてこれらを構成する各文を比較して、旧原文の各文の翻訳文は旧訳文のどの文に相当するか、その対応付けを行う部分である。対応付けを行うには、原文と訳文の文の類似度を調べ、対応付けを行った各文の類似度の和が最も大きくなるような組合せを選択すればよい。その具体例は後で説明するが、類似度を計算するには、原文の形態素解析を行い、単語単位に区切られた全ての語を単語辞書を用いて辞書引きをして、対応する訳文中にその語が存在するかどうかでポイント付けを行う。

【0030】単語辞書52は、ある語を入力すると、その訳語の候補を複数個返すような構成となっている。これは、一般的な英和辞典や和英辞典等に相当する。この単語辞書52は、機械翻訳処理部6で使用される単語辞書をそのまま流用することができる。

【0031】図11には、対応付け演算処理の具体的な内容を示す説明図を図示した。文の対応付けとは、ある

言語で記述された文書（原文）とそれを他の言語に翻訳した文書（訳文）がある場合に、原文中の文が訳文中のどの部分に対応しているかを自動的に判別することである。既に説明したように、各文書を1文ごとに分割し形態素解析を行って、単語辞書を用いて原文と翻訳文の類似度を計算する。この類似度の和が文書全体で最大になるような組合せを選択する。対訳文の対応組 (x, y) の評価関数 $h(x, y)$ には様々な式が考えられるが、例えば図11に示したような演算式を用いる。

【0032】ここで、 $f_m(x, y)$ は、文の組 x 中の自立語が文の組 y 中の自立語と一致するか、あるいは、 y 中の自立語を辞書引きしたものの中に存在する数を示す。 $f_j(x)$ は x 中に存在する自立語数である。 $\min(\dots)$ は原文と訳文の自立語数の比を計算しており、数が同じ場合が最大で1となるようにしてある。文の数の組合せは、 (x, y) が $(1, 1)$, $(1, 2)$, $(2, 1)$, $(1, 3)$, $(3, 1)$ の場合のいずれかがほとんどなので、この組合せの類似度を求めれば通常の文章では大抵うまくいく。このようにして行った対応付けの結果は文書対応付け部51が保存しておく。この自動対応付けで100%正解を見つけれられるわけではないので、対訳チェック表示部56に表示して、正しく対応付けが行われたかどうかをユーザが目でチェックすることができる。

【0033】図12には、対訳チェック表示部のウィンドウ説明図を示す。この図には、原文61と訳文62とが表示されている。スクロールバー63、64は原文61や訳文62の前後を比較して見るために操作する部分である。この例では、実際に1文対1文で対応付けが行われている。このユーザによるチェックが終了すると、終了キー65をクリックしてウィンドウを閉じる。

【0034】再び図10に戻って、この実施例では、先に図10で説明した処理選択指示部4によって差分値に応じた配分が行われる。即ち、差分値が閾値よりも大きい場合には機械翻訳処理部6に原文が送り込まれて機械翻訳を行う。その結果は翻訳チェック表示部54に表示されて、ユーザのチェックを受ける。また、差分値が閾値よりも小さい場合には差分表示部55に送られてやはりユーザによって内容が検討され、先に説明した要領で旧原文と相違部分の翻訳が行われる。また、旧原文と新原文とが一致する部分については、文書対応付け部51から訳文変換手段53に対応する訳文が送られ、そのまま対訳チェック表示部56に転送される。こうして、翻訳チェック表示部54や差分表示部55あるいは対訳チェック表示部56によって処理された結果は文書合成部9に転送され合成されて、新訳文が生成される。

【0035】図13には、図10に示した本発明の装置による処理フローチャートを示す。この図に示す多くの部分は、既に図9を用いて説明した通りのものである。図9と相違する点は、ステップS1において、旧原文フ

ァイル、新原文ファイルと共に旧訳文ファイルを読み込む点である。これによって、ステップS2で旧原文ファイルと旧訳文ファイルとの文の対応付けが行われる。この動作は、既に説明した通り、文書対応付け部51により行われる。その後、ステップS3において、文書比較部3が旧原文と新原文の差分を調べ、差分情報を作成する。その後のステップS4～ステップS11までの部分は、図9に示したステップS3～S10に該当する部分と同等の処理を行う。一方、この例では、ステップS6において、旧原文と新原文との間に差分が無いと判断された場合に、ステップS12において、旧訳文ファイルから原文に対応する訳文を見つけ、ステップS13で対応する原文及び訳文を対訳チェック表示部56に表示し、ユーザにチェックさせるようにしている。

【0036】以上の処理のうち、ステップS8やS11あるいはS13におけるユーザの修正作業は、1ブロックずつ対話的に行うこともできるし、またファイル全体の処理を行った後に一括して処理することもできる。ユーザはいずれの方法も選択が可能である。

【0037】本発明は以上の実施例に限定されない。上記実施例は英文から日本語への翻訳処理を例にとって説明したが、機械翻訳手段を交換すれば、日本語から英文への翻訳処理や他の言語への翻訳処理に対しても同様に実施することができる。もちろん、翻訳チェックや訳文入力時の表示画面は自由にレイアウト変更することが可能である。更に、差分率を計算する場合に計算対象となる範囲を行としたが、文や段落、ページ等の範囲で差分率を演算処理してもよい。また、単位データは単語でなく、1個1個の文字やあるいは行としても同様の効果が得られる。更に、旧原文と新原文の相違を定量的に表す方法として、差分率を求める方法の他、例えば機械翻訳部に元々備え付けられている構文解析処理部を利用し、構文解析の結果を比較して両者が似ているかどうかを判断するようにしてもよい。

【0038】

【発明の効果】以上説明した本発明の翻訳支援装置は、既に機械翻訳処理され、翻訳チェックされた旧原文の一部を変更した新原文について、これを翻訳しようとする場合に、旧原文と新原文の一致する部分は旧訳文をそのまま利用し、少し相違する部分は旧訳文の手直しにより新訳文を生成し、全く相違する部分は機械翻訳と翻訳チェック等を用いて翻訳処理することにより最終的な修正まで含めた翻訳時間を短縮し、翻訳コストを低減することができる。しかも、旧原文により実施した翻訳チェックによる旧訳文を十分効果的に利用することによって翻訳品質を維持し、あるいはその品質を一層高めることが可能になる。

【0039】また、旧原文と新原文とを比較したとき、差が全く無い文については、対応する旧訳文を取り出し、それをそっくりそのままコピーするようにすれば、

訳文の文書フォーマットもそのまま有効に利用し、重複する作業を防いで負担を軽減できる。従って、例えば旧原文と旧訳文とが完全に対応していないような場合でも、旧原文の一部について訳文が存在すれば、旧訳文の一部を利用して新原文の翻訳処理を効率的に実行することが可能となる。

【図面の簡単な説明】

【図1】本発明の翻訳支援装置実施例を示すブロック図である。

【図2】旧原文ファイルの例説明図である。

【図3】新原文ファイルの例説明図である。

【図4】旧訳文ファイルの例説明図である。

【図5】差分情報の例説明図である。

【図6】差分率の例説明図である。

【図7】翻訳チェック時の表示画面例説明図である。

【図8】訳文入力時の表示画面例説明図である。

【図9】本発明の装置による処理フローチャートである。

【図10】本発明の別の実施例の装置ブロック図である。

【図11】対応付け演算処理の説明図である。

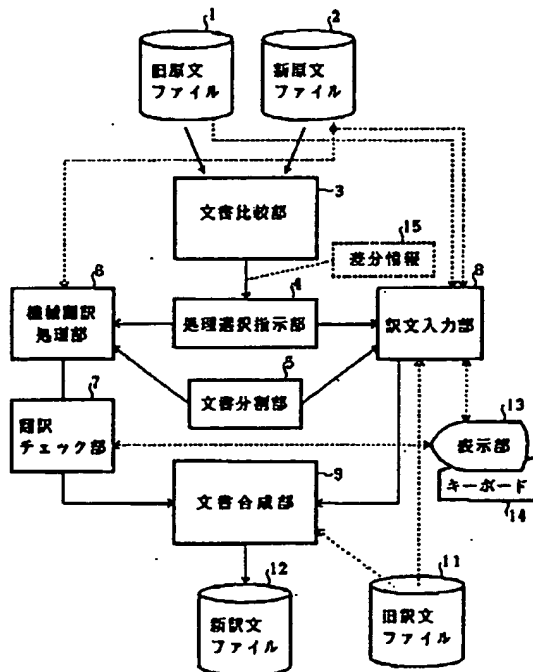
【図12】対訳チェック表示部のウインドウ説明図である。

【図13】本発明の別の装置による処理フローチャートである。

【符号の説明】

- 1 旧原文ファイル
- 2 新原文ファイル
- 3 文書比較部
- 4 処理選択指示部
- 5 文書分割部
- 6 機械翻訳処理部
- 7 翻訳チェック部
- 8 訳文入力部
- 9 文書合成部
- 10 旧訳文ファイル
- 11 新訳文ファイル
- 12 表示部
- 13 キーボード

【図1】



本発明の翻訳支援装置ブロック図

【図6】

行数	旧原文			新原文			差分率
	ワード数	差分ワード数	差分率	ワード数	差分ワード数	差分率	
1	14	1	7%	14	1	7%	7%
2	3	1	33%	3	1	33%	33%
3	9	3	33%	10	4	40%	40%

差分率の例説明図

【図11】

$$h(x, y) = \frac{f_m(x, y)}{f_j(x)} + \frac{f_m(y, x)}{f_j(y)} + \min\left(\frac{f_j(y)}{f_j(x)}, \frac{f_j(x)}{f_j(y)}\right)$$

対応付け演算処理の説明図

【図2】

1文→1行 単語→単位量のデータ

1	cat reads each filename in sequence and outputs it on the standard output. Thus:
2	exampleX cat goodies
3	outputs the contents of goodies on the standard output.
4	If no filename argument is given, or if the argument '-' is given, cat reads from the standard input. Beware of 'cat a b>a' and 'cat a b>b', which destroy the input files before reading them.

旧原文ファイルの例説明図

【図3】

1文→1行 単語→単位量のデータ

1	cat reads each filename in sequence and displays it on the standard output. Thus:
2	exampleX cat file1
3	displays the contents of file1 on the standard output, and
4	exampleX cat file1 file2>file3
5	concatenates the first two files and places the result on the third.
6	If no filename argument is given, or if the argument '-' is given, cat reads from the standard input. Beware of 'cat a b>a' and 'cat a b>b', which destroy the input files before reading them.

新原文ファイルの例説明図

【図4】

1	cat は、指定した各 filename を順次読みとり、標準出力に書き出します。例えば、
2	exampleX cat goodies
3	は、goodies の内容を標準出力に出力します。
4	入力ファイルを指定しない場合または引数に '-' を指定した場合は、cat は標準入力から読みとります。コマンド形式 'cat a b>a' および 'cat a b>b' は、読み取り前に入力ファイルのデータを破壊しますので注意して下さい。

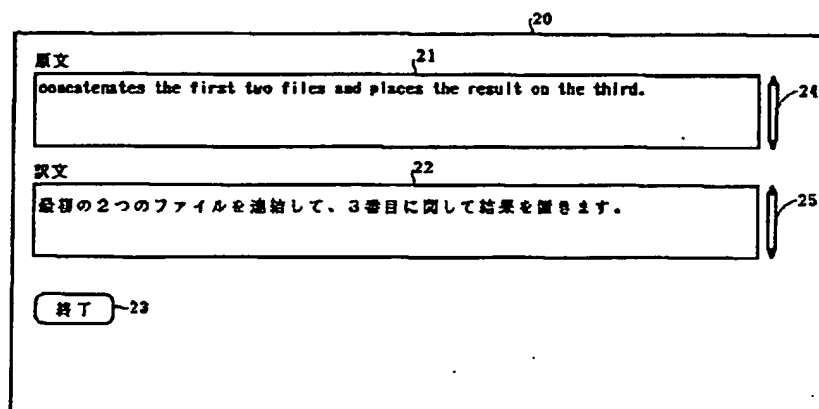
旧訳文ファイルの例説明図

【図5】

1	! cat reads each filename in sequence and [-outputs-][+displays+] it on the standard output. Thus:
2	! example% cat [-goodies-][+file!+]
3	! [-outputs-][+displays+] the contents of [-goodies-][+file!+] on the standard [-output.-][+output, and+]
4	+ example% cat file1 file2 > file3
5	+ concatenates the first two files and places the result on the third.
6	= If no filename argument is given, or if the argument '-' is given, cat reads from the standard input. Beware of 'cat a b > a' and 'cat a b > b', which destroy the input files before reading them.

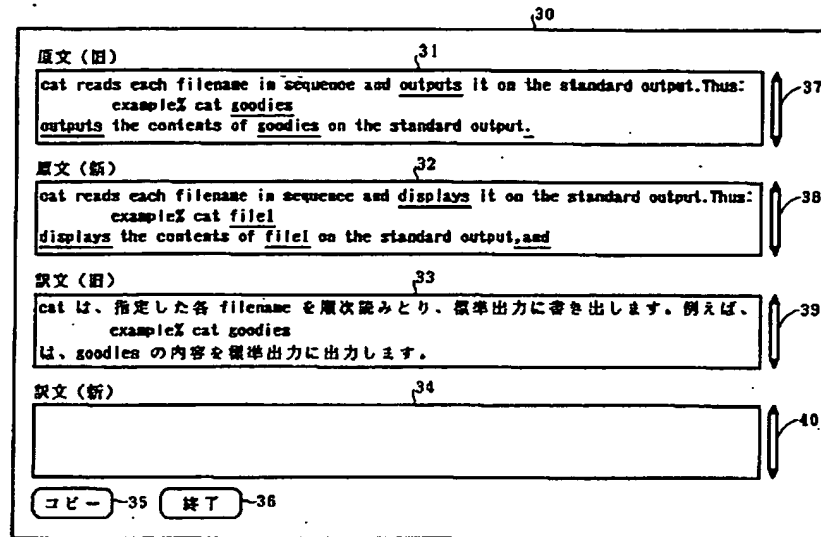
差分情報の例説明図

【図7】



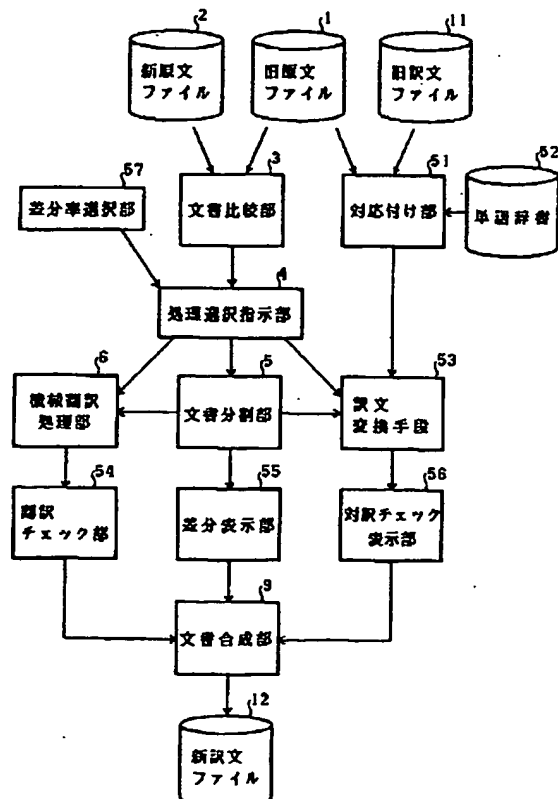
翻訳チェック時の表示画面例説明図

【図8】



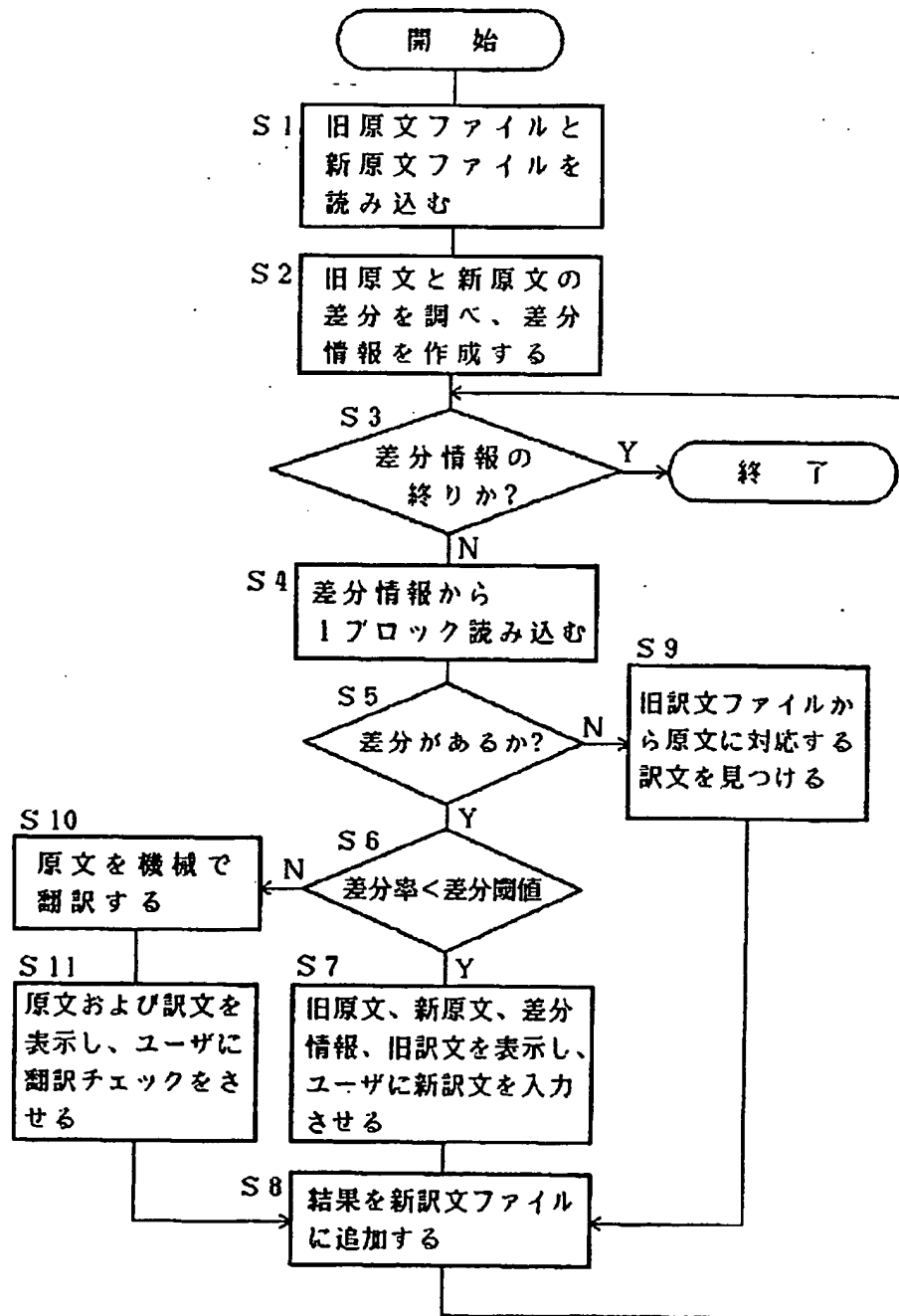
訳文入力時の表示画面説明図

【図10】



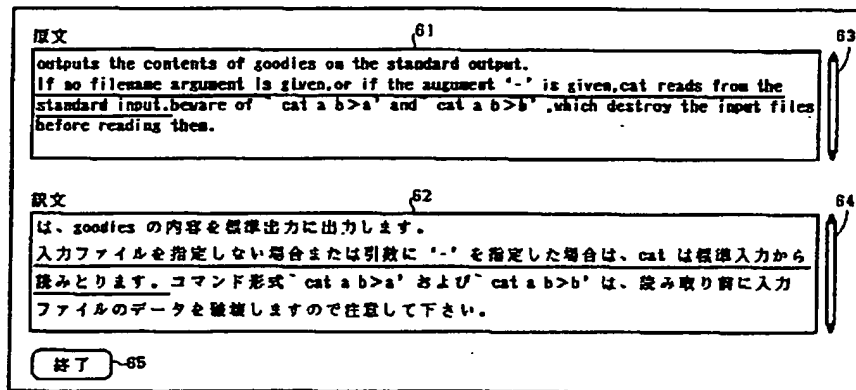
本発明の別の実施例の装置ブロック図

【図9】



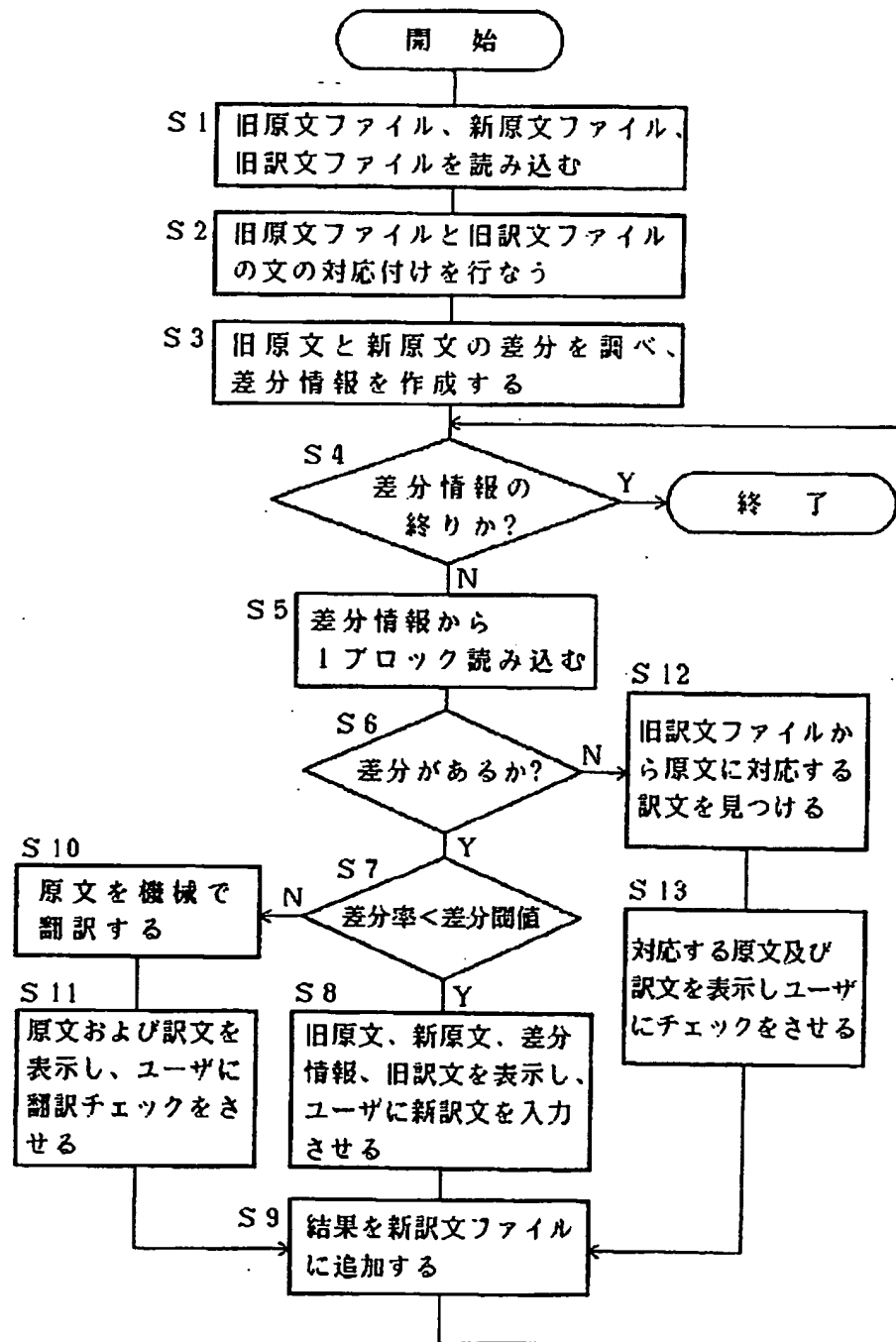
本発明の装置による処理フローチャート

【図12】



対訳チェック表示部のウィンドウ説明図

【図13】



本発明の別の装置による処理フローチャート